

**PERBANDINGAN 3 METODE DATA MINING UNTUK PENENTUAN HIPERTENSI
DI RUMAH SAKIT UMUM DAERAH Dr. MOEWARDI SURAKARTA**



**Disusun sebagai salah satu syarat menyelesaikan Jenjang Strata 1
pada Program Studi Informatika Fakultas Komunikasi dan Informatika**

Oleh:

TAUFIQ USMAN

L 200 120 054

**PROGRAM STUDI INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

2016

HALAMAN PERSETUJUAN

**PERBANDINGAN 3 METODE DATA MINING UNTUK PENENTUAN
HIPERTENSI DI RUMAH SAKIT UMUM DAERAH Dr. MOEWARDI
SURAKARTA**

PUBLIKASI ILMIAH

oleh:

TAUFIQ USMAN

L 200 120 054

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



Umi Fadlilah, S.T M.Eng.

NIP. 197803222005012002

HALAMAN PENGESAHAN

PERBANDINGAN 3 METODE DATA MINING UNTUK PENENTUAN
HIPERTENSI DI RUMAH SAKIT UMUM DAERAH Dr. MOEWARDI
SURAKARTA

OLEH

TAUFIQ USMAN

L 200 120 054

Telah dipertahankan di depan Dewan Penguji
Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta
Pada hari Sabtu, 19 Juni 2016
dan dinyatakan telah memenuhi syarat

Dewan Penguji:

1. Umi fadlilah, S.T M.Eng.
(Ketua Dewan Penguji)
2. Nurgiyatna, M.Sc., Ph.D
(Anggota I Dewan Penguji)
3. Yusuf Sulistyo Nugroho, S.T., M.Eng
(Anggota II Dewan Penguji)

(.....)
(.....)
(.....)

Publikasi ilmiah ini telah diterima sebagai salah satu persyaratan

Untuk memperoleh gelar sarjana

Tanggal 23 JUNI 2016

Mengetahui,

Dekan
Fakultas Komunikasi dan Informatika

Husni Thamrin, S.T., M.T., Ph.D.
NIK : 706

Ketua Program Studi
Informatika

Dr. Heru Supriyono, M.Sc.
NIK:970

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 17 Juni 2016

Penulis



TAUFIQ USMAN

L 200 120 054

PERBANDINGAN 3 METODE DATA MINING UNTUK PENENTUAN HIPERTENSI DI RUMAH SAKIT UMUM DAERAH Dr. MOEWARDI SURAKARTA

Abstrak

Perubahan gaya hidup dan banyaknya unsur makanan mengandung lemak tinggi serta minuman yang mengandung alkohol yang dikonsumsi oleh masyarakat dapat menimbulkan terjadinya hipertensi, hal ini bisa terjadi karena masyarakat kurang mengetahui faktor-faktor hipertensi. Dengan demikian data pasien di RSUD Dr. Moewardi surakarta yang sebelumnya hanya dibiarkan saja dan hanya menimbulkan penumpukan data yang besar, hal ini bisa dijadikan informasi yang berguna untuk penentuan hipertensi. Penelitian ini bertujuan untuk penentuan hipertensi berdasarkan faktor-faktor yang paling mempengaruhi menggunakan teknik data mining dengan menggunakan tiga metode yaitu *decision tree* Algoritma ID3, *naïve bayes* dan *regresi linier*. Adapun atribut yang digunakan adalah Jenis Kelamin, Kategori Usia, Asupan Lemak dan Konsumsi Alkohol. Hasil perbandingan metode ini menggunakan Software *Rapid Miner 5* dan *Microsoft Excel* untuk mengetahui metode yang paling akurat, sehingga dari hasil implementasi tiga metode dapat diketahui berdasarkan nilai *accuracy*, *precision*, *recall* metode regresi linier lebih akurat digunakan dengan nilai *accuracy* 72,09% dan *precision* 78,52%. Atribut yang paling mempengaruhi dalam penentuan hipertensi di RSUD Dr. Moewardi Surakarta adalah Konsumsi Alkohol.

Kata Kunci : *Data Mining, Decision Tree, Algoritma ID3, Hipertensi, Naive Bayes, Regresi Linier*

Abstract

changes of lifestyle and many foods containing fat high as well as drinks containing alcohol that consumed by the community could cause the occurrence of hypertension, it can happen because the community doesn't know factor of hypertension. This data patients in RSUD Dr. Moewardi Surakarta who previously only will be left alone and only causing accumulation of big data, this could be useful information to determine the hypertension. This research aims to the determination of hypertension based on factor the most affect using in data mining techniques with using three method are *Decision tree* Algoritma ID3, *naïve bayes* and *regresi linier*. While attribute – an attribute that used is gender, age categories, intake fat, consumption of alcohol and the results of hypertension. The results of a comparison this method using Software *Rapidminer 5* and *Microsoft Excel* to know a method of the most accurate, so from the implementation of three method it can be seen based on virtue of value *accuracy*, *precision*, *recall*. Method of linear regression it is better used with having value *accuracy* 72,09% and *precision* 78,52% . An attribute that most affect in the determination of hypertension in RSUD Dr. Moewardi Surakarta is the consumption of alcohol.

Kata Kunci : *Data Mining, Decision Tree, ID3, Hypertension, Naïve Bayes*

1. PENDAHULUAN

Rumah Sakit Umum Daerah Dr. Surakarta memberikan pelayanan untuk pemeriksaan tekanan darah di poli dalam, adapun penyakit hipertensi yang diklasifikasikan menjadi dua kategori yaitu YA dan TIDAK . Hipertensi adalah keadaan dimana *supply* oksigen dan nutrisi yang dibawa oleh darah terhambat dalam proses pengirimannya ke jaringan tubuh yang membutuhkan. Penyakit ini dapat menimbulkan gangguan pada organ tubuh lain karena menyebabkan organ-organ tersebut harus bekerja lebih keras, keadaan ini disebut dengan komplikasi. (Ningrat Dkk, 2012). Penelitian ini menerapkan *data mining* dengan membandingkan tiga metode dengan tujuan untuk memudahkan mencari faktor-faktor yang paling mempengaruhi dan mencari algoritma yang paling akurat dalam penentuan hipertensi serta memanfaatkan data yang setiap harinya meningkat dan hanya dibiarkan menumpuk, hal ini bisa dijadikan informasi yang bermanfaat dalam penentuan hipertensi.

Data mining adalah proses yang menggunakan teknik statistik, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar. (Ridwan Dkk, 2013). *Decision tree* dengan algoritma ID3 adalah metode yang menggunakan pohon keputusan dan konsep dari *entropy* informasi dalam pemilihan atau penentuan atribut dengan menggunakan *Information Gain*. (Obbie, 2015). *Naïve bayes* adalah suatu metode klasifikasi atau probabilitas yang berdasarkan teorema bayes, naïve bayes digunakan untuk membuat model dengan kemampuan prediktif pada studi kasus. (Manjusha et al, 2014). Regresi linear adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (dependen) dengan satu atau lebih variabel bebas (independen). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, (Susanto, 2014). Berdasarkan permasalahan tersebut maka dalam penelitian ini akan dilakukan perbandingan terhadap 3 metode untuk menentukan hipertensi menggunakan *data mining*, yaitu metode *Decision Tree* algoritma ID3, *Naive Bayes* dan *Regresi Linier*. Diharapkan setelah diterapkan *data mining*, dapat memberikan informasi penentuan hipertensi dengan mudah dan mengetahui faktor-faktor yang paling mempengaruhi, serta mencari algoritma dengan nilai akurasi yang tinggi dengan hasil rekomendasi terbaik.

2.METODE PENELITIAN

2.1 Penentuan Atribut dan Variabel

Atribut dan variabel yang digunakan dalam *data mining* ini dihasilkan berdasarkan dari tujuan penelitian dan konsultasi kepada dokter di poli dalam, atribut dan variabel yang digunakan yaitu seperti pada Tabel 1 :

Tabel 1. Atribut dan Variabel Pasien

Atribut	Variabel
Jenis Kelamin	X1
Kategori Usia	X2
Asupan Lemak	X3
Konsumsi Alkohol	X4
Hasil Hipertensi	Y

Atribut dan variabel di hasilkan dari hasil konsultasi dengan dokter di poli dalam tentang faktor-faktor yang mempengaruhi hipertensi di RSUD Dr.Moewardi surakarta.

2.2 Pengumpulan Data Pasien

Data yang digunakan adalah data pasien pemeriksaan hipertensi di poli dalam dan rekam medik tahun 2015 di RSUD Dr. Moewardi Surakarta.

2.3 Penentuan Sampel Data Pasien

Jumlah data pasien pada tahun 2015 yang diperoleh dari poli dalam dan rekam medik di rumah sakit Dr. Moewardi surakarta dihitung dengan metode *Slovin* untuk menentukan jumlah sampel yang digunakan sebagai data pengujian proses *data mining* (Nugroho, 2014) seperti pada **persamaan 1.**

$$n = \frac{N}{1+Ne^2} \dots\dots\dots (1)$$

Dimana :

n = jumlah Sampel

N = jumlah keseluruhan data

e = persen toleransi ketidaktelitian dapat ditolerir (max 5%)

2.4 Penerapan *Naive Bayes*

Naïve Bayes adalah suatu metode klasifikasi atau probabilitas yang berdasarkan teorema bayes, naïve bayes digunakan untuk membuat model dengan kemampuan prediktif pada studi kasus.(Manjusha et al, 2014)

Persamaan 2 teorema Bayes :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots(2)$$

2.5 Penerapan *Decision Tree* Algoritma ID3

Decision tree adalah salah satu metode algoritma yang mempunyai pohon keputusan untuk menguji setiap node, serta menggunakan *entropy* dan *information gain* dalam menghitung setiap atribut.(Lakshmi et al, 2014)

Persamaan 3 Rumus *Entropy* :

$$Entropi(y) = -p_1\log_2p_1 - p_2\log_2p_2 - \dots - p_n\log_2p_n \dots\dots\dots(3)$$

Persamaan 4 Rumus *Information Gain* :

$$gain(y, A) = entropi(y) - \sum_{c \in nilai(A)} \frac{y_c}{y} entropi(y_c) \dots\dots\dots(4)$$

2.6 Penerapan Regresi Linier

Regrsi linier adalah suatu metode statisitk yang sangat penting yang menyelidiki hubungan antara variabel respon Y (independen) dan variabel X (Prediktor). (Smadi & Abu-Afouna, 2012)

Persamaan 5 Regresi Linier :

$$Y_i = b_0 + b_1x_1 \dots\dots\dots(5)$$

3.HASIL DAN PEMBAHASAN

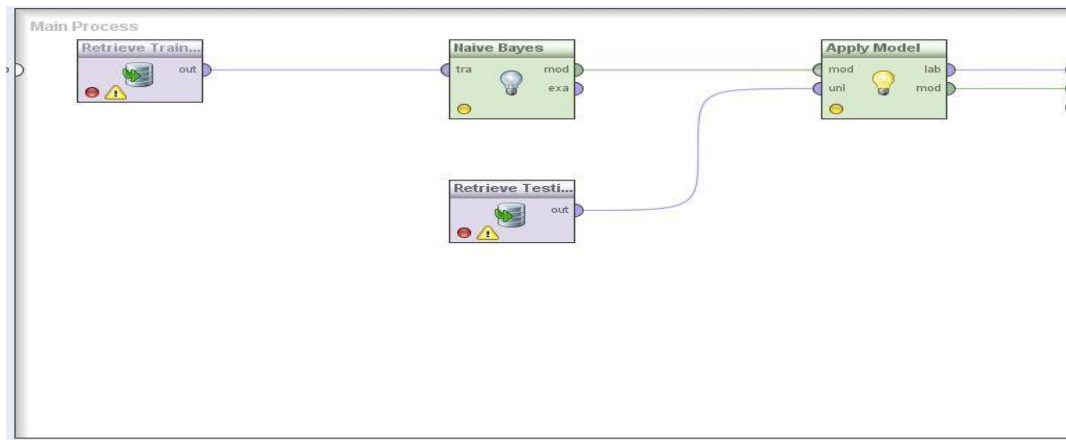
3.1 Pengambilan Sampel Data Pasien

Diketahui dalam satu tahun jumlah pasien pemeriksaan di poli dalam RSUD Dr.Moewardi Surakarta memiliki jumlah 7400 pasien dan memiliki toleransi ketidaktelitian 5%, Maka jumlah sampel yang akan diambil bisa dihitung berdasarkan **Persamaan 1**.

$$\begin{aligned} n &= 7400 / 1 + 7400 \times (0.05)^2 \\ n &= 7400 / 1 + 7400 \times 0.0025 \\ n &= 7400 / 1 + 18.5 \\ n &= 7400 / 19.5 \\ n &= 379,487179487 \text{ pasien} \\ \text{Jadi dibulatkan menjadi } &379 \text{ pasien.} \end{aligned}$$

3.2 Implementasi *Naive Bayes* Menggunakan Software *RapidMiner 5*

Rancangan Proses untuk memprediksi data pasien Hipertensi menggunakan software *RapidMiner 5* seperti **Gambar 1**.



Gambar 1. Rancangan Proses *Naive Bayes* menggunakan *data training*.

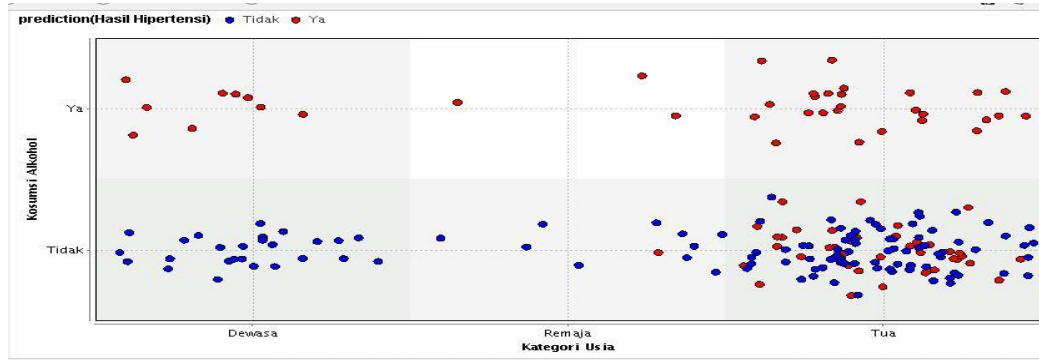
Hasil prediksi data pasien yang sudah di proses dapat dilihat pada Gambar 2.

ExampleSet (195 examples, 3 special attributes, 4 regular attributes)								View Filter (195 / 195): a	
Row No.	confidence(Tidak)	confidence(Ya)	prediction(Hasil Hiperten...	jenis kelamin	Kategori Usia	Asupan Lemak	Konsumsi Alkohol		
1	0.891	0.109	Tidak	Perempuan	Dewasa	Cukup	Tidak		
2	1.000	0.000	Tidak	Laki-laki	Remaja	Cukup	Tidak		
3	0.050	0.950	Ya	Laki-laki	Tua	Lebih	Ya		
4	0.891	0.109	Tidak	Perempuan	Dewasa	Cukup	Tidak		
5	0.891	0.109	Tidak	Perempuan	Dewasa	Cukup	Tidak		
6	0.891	0.109	Tidak	Perempuan	Dewasa	Cukup	Tidak		
7	0.647	0.353	Tidak	Laki-laki	Dewasa	Lebih	Tidak		
8	0.891	0.109	Tidak	Perempuan	Dewasa	Cukup	Tidak		
9	0.846	0.154	Tidak	Laki-laki	Dewasa	Cukup	Tidak		
10	0.846	0.154	Tidak	Laki-laki	Dewasa	Cukup	Tidak		
11	0.732	0.268	Tidak	Perempuan	Dewasa	Lebih	Tidak		
12	0.891	0.109	Tidak	Perempuan	Dewasa	Cukup	Tidak		
13	0.386	0.614	Ya	Perempuan	Tua	Lebih	Tidak		
14	0.653	0.347	Tidak	Perempuan	Tua	Cukup	Tidak		
15	0.050	0.950	Ya	Laki-laki	Tua	Lebih	Ya		
16	0.406	0.594	Ya	Laki-laki	Dewasa	Cukup	Ya		
17	0.135	0.865	Ya	Laki-laki	Tua	Cukup	Ya		
18	0.999	0.001	Tidak	Laki-laki	Remaja	Lebih	Tidak		

Gambar 2. Tampilan hasil *Naive Bayes* untuk memprediksi hasil Hipertensi pada *Result Perspective*

Berdasarkan Gambar 2 menunjukkan hasil prediksi penentuan hipertensi dengan nilai *confidence* “tidak” dan nilai *confidence* “ya”. Dengan jenis kelamin “perempuan” dan kategori usia “dewasa”, dengan asupan lemak “cukup” dan konsumsi alkohol “tidak” menghasilkan nilai *confidence* “tidak” 0,891 dan nilai *confidence* “ya” 0,109 serta menghasilkan prediksi hasil hipertensi “Tidak”.

Hasil pada *Scatter Plot Naive Bayes* berdasarkan data testing dapat dilihat pada Gambar 3.

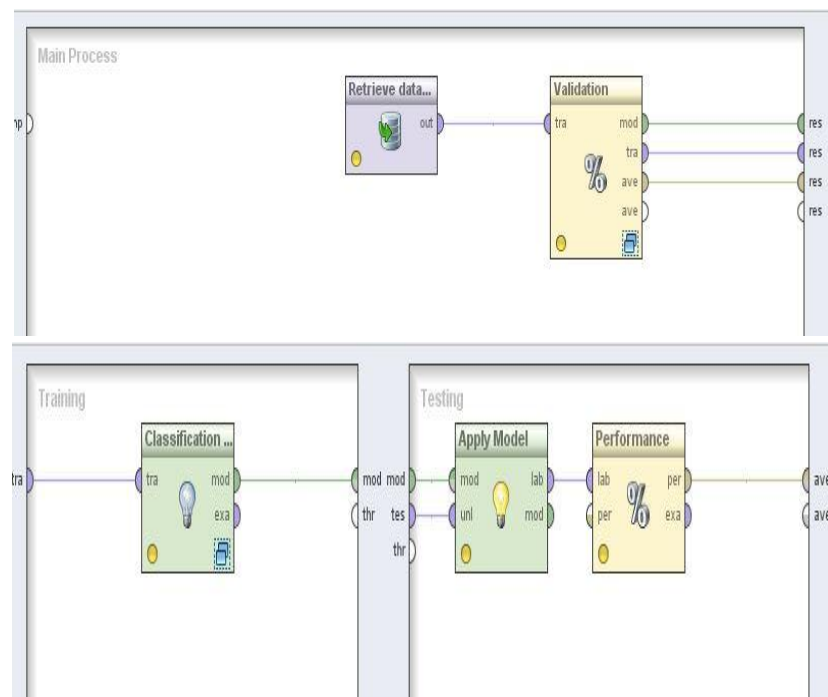


Gambar 3. Tampilan Hasil Naive Bayes Scatter Plot Menggunakan Data Training

Berdasarkan *plot view* pada gambar 3, menunjukkan penentuan hipertensi pasien dengan konsumsi alkohol “Ya” dengan kategori usia “Dewasa”, “remaja”, “tua” menghasilkan lebih dominan “Ya” hipertensi, sedangkan untuk konsumsi alkohol “tidak” dengan kategori usia “dewasa”, “remaja”, “tua” menghasilkan lebih dominan “Tidak” hipertensi. Sedangkan konsumsi alkohol “ya” dengan kategori usia “remaja” menghasilkan “Ya” hipertensi. Sedangkan untuk konsumsi alkohol “tidak” dengan kategori usia “remaja” menghasilkan “Tidak” hipertensi.

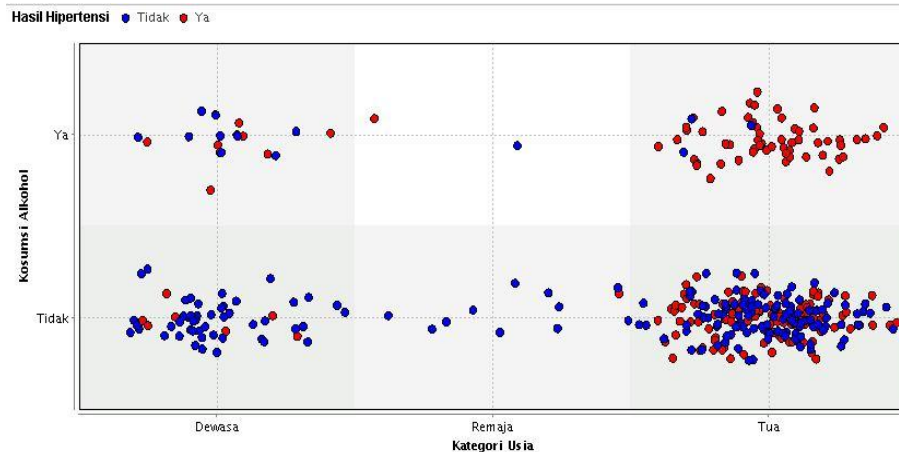
3.3 Implementasi *Regresi Linier* menggunakan Software *RapidMiner 5*

Rancangan proses dengan metode *Regresi Linier* menggunakan Software *RapidMiner 5* ditunjukkan pada Gambar 4.



Gambar 4. Rancangan Regresi Linier menggunakan Software RapidMiner 5

Tampilan untuk melihat hasil dari *Scatter Plot*, hasil dapat dilihat pada Gambar 5.

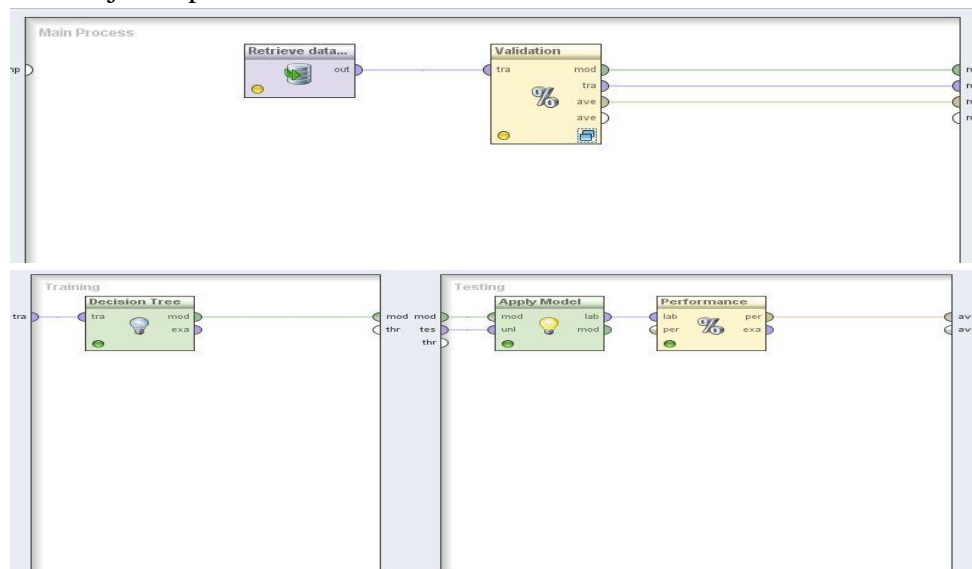


Gambar 5. Hasil dari *Scatter Plot* menggunakan Software *RapidMiner 5*

Berdasarkan *plot view* pada gambar 5 menunjukkan penentuan hipertensi pada pasien dengan konsumsi alkohol “ya” dan kategori usia “dewasa”, “tua” menghasilkan lebih dominan “Ya” hipertensi. Sedangkan untuk konsumsi alkohol “tidak” dan kategori usia “tua”, “dewasa” menghasilkan lebih dominan “Tidak” hipertensi. Untuk konsumsi alkohol “ya”, “tidak” dengan kategori usia “remaja” menghasilkan “Ya” hipertensi.

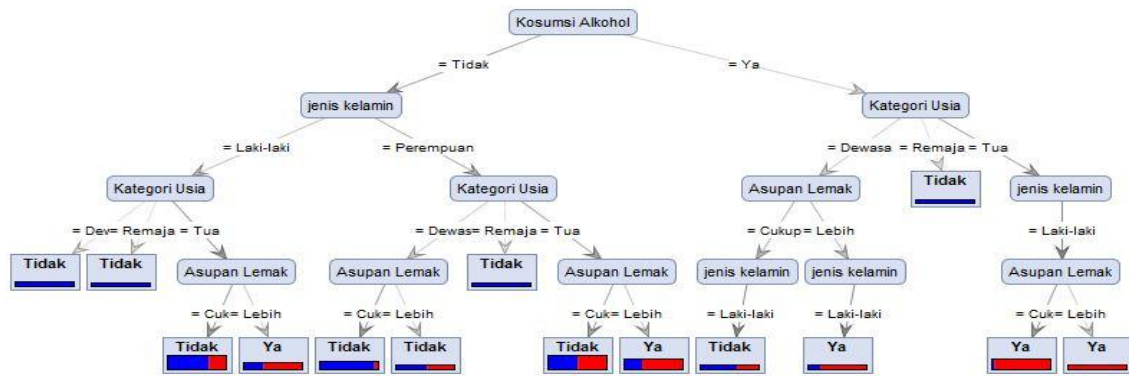
3.4 Implementasi *Decision Tree* menggunakan Software *RapidMiner 5*

Rancangan proses algoritma klasifikasi data pasien Hipertensi menggunakan Software *RapidMiner 5* ditujukan pada Gambar 6.



Gambar 6. Rancangan proses *Decision Tree* menggunakan data training

Hasil skema pohon yang telah di proses data pasien pemeriksaan hipertensi menggunakan Software *RapidMiner 5* dapat dilihat pada Gambar 7



Gambar 7. Hasil tampilan skema pohon Decision Tree untuk klasifikasi penentuan hipertensi

Dari hasil skema pohon keputusan pada Gambar 7 atribut konsumsi alkohol menempati posisi sebagai *root node*. sedangkan posisi *internal node* pertama pada konsumsi alkohol “ya” adalah kategori usia, sedangkan posisi *internal node* pertama pada konsumsi alkohol “tidak” adalah jenis kelamin. Semua atribut yang digunakan dalam penelitian ini mempengaruhi dalam penentuan hipertensi pada pasien.

3.5 Implementasi Perhitungan *Naive Bayes* Menggunakan Software Microsoft Excel

Sebagai contoh perhitungan diambil data pasien yang memiliki ciri sebagai berikut : Jenis Kelamin “perempuan” Kategori Usia “tua” Asupan Lemak “cukup” dan Konsumsi Alkohol “tidak”. Apakah pasien tersebut mempunyai gejala Ya hipertensi atau Tidak Hipertensi ? hasil perhitungan menunjukkan **Persamaan 2**.

$$P(Y = Ya) = 187/379 = 0,493$$

$$P(Y = Tidak) = 192/379 = 0,507$$

$$P(X_1 = Perempuan | Ya) = 76/187 = 0,406$$

$$P(X_1 = Perempuan | Tidak) = 97/192 = 0,505$$

$$P(X_2 = Tua | Ya) = 169/187 = 0,904$$

$$P(X_2 = Tua | Tidak) = 125/192 = 0,651$$

$$P(X_3 = Cukup | Ya) = 109/187 = 0,582$$

$$P(X_3 = Cukup | Tidak) = 155/192 = 0,807$$

$$P(X_4 = Ya | Ya) = 118/187 = 0,631$$

$$P(X_4 = Ya | Tidak) = 179/192 = 0,932$$

HMAP dari fakta di atas dapat dihitung dengan :

$$P(X_1 = Perempuan, X_2 = Tua, X_3 = Cukup, X_4 = Ya | Y = Ya) = \frac{76}{187} \times \frac{169}{187} \times \frac{109}{187} \times \frac{118}{187} \times \frac{187}{379} = 0,067$$

$$P(X_1 = Perempuan, X_2 = Tua, X_3 = Cukup, X_4 = Tidak | Y = Tidak) = \frac{97}{192} \times \frac{125}{192} \times \frac{155}{192} \times \frac{179}{192} \times \frac{192}{379} = 0,125$$

KEPUTUSAN PENENTUAN HIPERTENSI = **TIDAK HIPERTENSI**.

3.6 Implementasi Perhitungan *Regresi Linear* Menggunakan Software Microsoft Excel

Pada metode regresi rumus dihasilkan dari rumus *rapidminer* 5 sebagai contoh diambilkan satu data uji sebagai berikut : $Y = 0,309 * (\text{Jenis Kelamin}) - 0,400 * (\text{Kategori Usia}) - 0,512 * (\text{Asupan Lemak}) - 1,122 * (\text{Konsumsi Alkohol}) + 0,874$ $- 0,309 * (\text{Jenis Kelamin}) + 0,400 * (\text{Kategori Usia}) + 0,512 * (\text{Asupan Lemak}) + 1,122 * (\text{Konsumsi Alkohol}) - 0,874$. Memasukkan nilai variabel X ke dalam model regresi :

$X_1= 1, X_2=1, X_3=0, X_4=0$. Sehingga, $Y = (0,309*1) - (0,400*1) - (0,512*0) - (1,122*0) + 0,874 - (0,309*1) + (0,400*1) + (0,512*0) + (1,122*0) - 0,874 = 0$.

Hasil perhitungan dengan menggunakan rumus hasil *rapidminer* 5 menghasilkan nilai nol, yang berarti “Tidak” hipertensi.

3.7 Implementasi Perhitungan *Decision Tree* Menggunakan Software Microsoft Excel

3.7.1 Menentukan *Root Node* Atribut

Atribut yang memiliki nilai *information Gain* tertinggi dipilih sebagai *Root Node*. Hasilnya dapat dilihat pada Tabel 3.

Tabel 3. *Information Gain* tertinggi

Atribut	Nilai <i>gain</i>
Jenis Kelamin	0,007
Kategori Usia	0,078
Asupan Lemak	0,044
Konsumsi Alkohol	0,104

Dari Tabel 3 dapat disimpulkan bahwa atribut konsumsi alkohol menempati sebagai *root note* dapat dilihat pada Gambar 8.



Gambar 8. *Root Node*

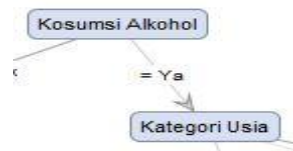
3.7.2 Menentukan *Internal Node* pertama Atribut

Menentukan *Internal Node* pada Konsumsi Alkohol “Ya” didapatkan nilai *Informatin Gain* pada Tabel 4.

Tabel 4. Nilai *Informaton Gain* Kategori Usia

Nilai <i>Information gain</i>	Konsumsi Alkohol “Ya”
Jenis Kelamin	0,000
Kategori Usia	0,178
Asupan Lemak	0,012

Dari hasil Tabel 4 dapat disimpulkan bahwa atribut Kategori Usia merupakan *internal node* pertama pada konsumsi alkohol “ya”. Dapat dilihat pada Gambar 9.



Gambar 9. Internal Node Pertama

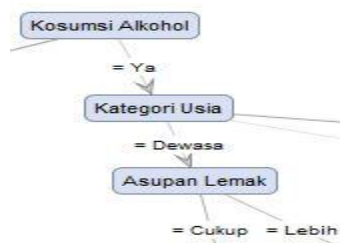
3.7.3 Menentukan *Internal node* kedua Atribut

Menentukan *Internal Node* pada konsumsi alkohol “ya” dengan kategori usia “dewasa” hasilnya seperti pada Tabel 5.

Tabel 5. Nilai *Information Gain* konsumsi alkohol “ya” dengan kategori usia “Dewasa”

Nilai <i>Information Gain</i>	Konsumsi Alkohol “ya” dengan Kategori Usia “Dewasa”
Jenis Kelamin	0,000
Asupan Lemak	0,129

Dari hasil Tabel 5 dapat disimpulkan bahwa atribut asupan lemak merupakan *internal node* kedua pada konsumsi alkohol “ya” dengan kategori usia “dewasa” Dapat dilihat pada Gambar 10.



Gambar 10. Internal Node Kedua

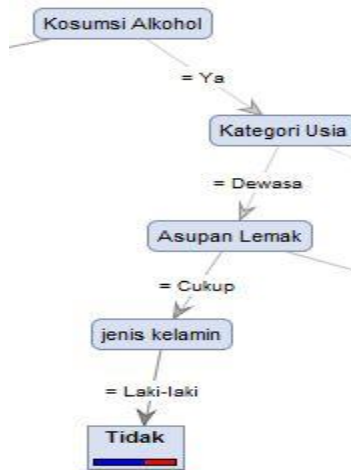
3.7.4 Menentukan *Internal Node* ketiga dan *leaf node* Atribut

Menentukan *internal node* ketiga dan *leaf node* pada konsumsi alkohol “ya” dengan kategori usia “dewasa” pada asupan lemak “cukup” dan jenis kelamin “laki-laki” hasilnya seperti pada Tabel 6.

Tabel 6. Nilai *Information Gain* Jenis Kelamin “laki-laki”

Nilai <i>Information Gain</i>	Konsumsi Alkohol “ya” , Kategori Usia “Dewasa” Asupan Lemak “cukup” dan Jenis Kelamin “laki-laki”
Jenis Kelamin	0,000

Dari hasil Tabel 6 dapat disimpulkan bahwa konsumsi alkohol “ya” dengan kategori usia “dewasa” asupan lemak “cukup” dan jenis kelamin “laki-laki” menghasilkan *leaf node* “Tidak” Hipertensi dikarenakan hasil dari *Information Gain* bernilai 0 dapat dilihat pada Gambar 11.



Gambar 11. Internal Node ketiga dan Leaf Node

3.8 Hasil Perbandingan 3 Metode Data Mining

Setelah melakukan analisa dari 3 metode tersebut untuk mengetahui metode tinggi tingkat akurasi dengan hasil rekomendasi yang paling baik dalam penelitian ini yaitu *Decision Tree*, *Naive Bayes* dan *Regresi Linier*. Maka selanjutnya melakukan perbandingan berdasarkan nilai *accuracy*, *Precision* dan *Recall*. Hasil perbandingan dapat dilihat pada Tabel 7.

Tabel 7. Hasil Perbandingan 3 Metode Data Mining

Komponen	<i>Decision Tree</i>	<i>Naive Bayes</i>	<i>Regresi Linier</i>
<i>Accuracy</i>	70,98 %	72,56 %	73,09 %
<i>Precision</i>	73,91 %	76,10 %	78,52 %
<i>Recall</i>	63,64 %	64,71 %	62,57 %

Berdasarkan hasil perbandingan pada Tabel 7, dapat disimpulkan bahwa metode *Regresi Linier* merupakan metode yang memiliki tingkat *Accuracy* 73,09% dan *Precision* 78,52% lebih baik digunakan dalam penelitian ini, sedangkan metode *Naive Bayes* memiliki tingkat *Recall* 64,71% yang lebih baik, dan metode *Decision Tree* memiliki tingkat *Recall* 63,64% lebih baik dibandingkan dengan metode *Regresi Linier* 62,57%.

3.9 Hasil Penelitian 3 Metode *Data Mining*

Dari hasil analisa yang dilakukan memberikan gambaran dalam penentuan Hipertensi dengan tiga metode. Metode yang digunakan adalah metode *Decision Tree* Algoritma ID3, *Naive Bayes* dan *Regresi Linear*. Hasil penelitian ini dapat diketahui sebagai berikut:

1. Hasil penelitian menggunakan metode *Decision Tree* Algoritma ID3

Dalam penelitian ini menghasilkan Konsumsi Alkohol sebagai atribut yang paling mempengaruhi. Hal Ini dibuktikan pada metode *Decision Tree* bahwa Konsumsi Alkohol yang menempati posisi sebagai *root node*. Aturan yang terbentuk dari hasil penelitian bahwa jika konsumsi alkohol “ya” maka tergantung pada kategori usia.

- a. Pada konsumsi alkohol “ya” dengan kategori usia “dewasa” dipengaruhi asupan lemak “cukup” dan dengan jenis kelamin “laki-laki” maka pasien menderita “Ya” hipertensi,
- b. Untuk konsumsi alkohol “ya” dengan kategori usia “dewasa” dipengaruhi asupan lemak “lebih” dan dengan jenis kelamin “laki-laki” maka pasien menderita “ya” hipertensi.
- c. Untuk konsumsi alkohol “ya” dengan kategori usia “remaja” maka pasien “tidak” hipertensi. Sedangkan untuk konsumsi alkohol “ya” dengan jenis kelamin “laki-laki” dan dipengaruhi asupan lemak “cukup” maka pasien menderita “ya” hipertensi.
- d. Untuk konsumsi alkohol “ya” kategori usia “tua” dengan jenis kelamin “laki-laki” dan dipengaruhi asupan lemak “lebih” maka pasien menderita “ya” hipertensi.
- e. Pada konsumsi alkohol “tidak” dapat dipastikan akan menghasilkan pasien “tidak” hipertensi, kecuali pasien dengan kategori usia “tua” dan dipengaruhi asupan lemak “lebih” maka pasien penderita “ya” hipertensi.

2. Hasil penelitian menggunakan metode *Naive Bayes*

- a. Dalam penelitian ini menghasilkan prediksi yaitu pada pasien dengan ciri-ciri jenis kelamin “perempuan” dengan kategori usia “tua” asupan lemak “cukup” dan konsumsi alkohol “tidak” menghasilkan pasien “tidak” hipertensi.
- b. Sedangkan pada pasien dengan ciri-ciri jenis kelamin “laki-laki” dengan kategori usia “tua” asupan lemak “lebih” dan konsumsi alkohol “ya” menghasilkan pasien “ya” hipertensi.

3. Hasil penelitian menggunakan metode *Regresi Linier*

Dalam penelitian ini menghasilkan rumus yang dihasilkan oleh *rapidminer 5* untuk penentuan hipertensi dapat diambil contoh pada penelitian ini :

- a. Pada pasien dengan ciri-ciri jenis kelamin “laki-laki” dengan kategori usia “dewasa” dan asupan lemak “cukup” konsumsi alkohol “tidak” dan menghasilkan pasien “Tidak” hipertensi
- b. Sedangkan pada pasien dengan ciri-ciri jenis kelamin “laki-laki” dengan kategori usia “tua” dan asupan lemak “lebih” konsumsi alkohol “ya” menghasilkan pasien “Ya” hipertensi.

4.PENUTUP

Berdasarkan hasil penelitian, maka dapat disimpulkan bahwa :

- a. Konsumsi Alkohol merupakan atribut yang paling mempengaruhi dalam penelitian ini. Hal ini dibuktikan pada metode *Decision Tree* yang menghasilkan konsumsi alkohol menempati sebagai *root node*.
- b. Berdasarkan nilai *accuracy* dan *precision*, metode *Regresi Linier* memiliki nilai yang lebih baik dibandingkan dengan metode yang lain, dengan nilai *accuracy* 73,09% dan *precision* 78,52%.
- c. Berdasarkan nilai *recall* metode *Naive Bayes* memiliki nilai lebih tinggi dibandingkan dengan metode yang lain, dengan nilai *recall* 64,71%.
- d. Data yang melimpah dan menumpuk yang hanya menjadikan beban *database* bisa di jadikan informasi yang berguna dan bermanfaat dengan menerapkan *data mining*.

PERSANTUNAN

Terima kasih untuk dosen pembimbing ibu Umi Fadlilah, S.T.M,Eng, dan dosen Yusuf Sulistyo Nugroho S.T.,M,Eng, Prof.Dr. Bambang Purwanto, Sp.PD-KGH., FINASIM, Dr.Eva Niamuziswati, Sp.PD, M.Kes dan pihak-pihak yang telah membantu untuk penelitian skripsi dan penulisan artikel ilmiah.

DAFTAR PUSTAKA

- Lakshmi,D.Bhu.; Arundathi,S.; DR.Jagadeesh.2014.”*A prediction for Student's Performance Using Decision Tree ID3 Method*”. International Journal of Scientific & Engineering Research, Volume 5, Issue 7, 1329 ISSN 2229-5518.
- Manjusha, K,K.; Sankaranarayanan,K.; P,seena.2014 “*Prediction of Different Dermatological Using Naïve Bayesian Classification*”. International Journal of Advanced Research In Computer Science and Software Engineering. Volume 4, Issue 1, ISSN: 2277 128X.
- Ningrat, Ranny Wahyu.; Santosa, Budi. 2012.”*Pemilihan Diet Nutrien bagi Penderita Hipertensi Menggunakan Metode Klasifikasi Decision Tree*”. Jurnal Teknik ITS, Vol. 1, No.1. ISSN: 2301-9271.
- Nugroho, Yusuf Sulistyo. 2014.”*Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta*”. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Yogyakarta, ISSN: 1979-911X.
- Obbie, Kristanto. 2015. “ *Penerapan Algoritma Klasifikasi Data Mining ID3 untuk Menentukan Penjurusan Siswa SMAN 6 Semarang* “. Jurusan Teknik Informatika.
- Ridwan, Mujib.; Suyono, Hadi.; Sarosa, M. 2013. “*Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*”. Jurnal EECCIS Vol.7, No.1.

- Smadi, Abdullah A.; Abu-Afouna.H, Nour. 2012. "*On Least Squares Estimation in a Simple Linear Regression Model with Periodically Correlated Errors: A Cautionary Note*". Austrian Journal of Statistics, Vol. 41, No. 3, 211–226.
- Susanto, Heri. 2014. "*Data Mining untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu*". Jurnal Pendidikan Vokasi, Vol 4, Nomor 2. Universitas Negeri Yogyakarta.



PEMERINTAH PROVINSI JAWA TENGAH
RUMAH SAKIT UMUM DAERAH Dr. MOEWARDI

Jl. Kolonel Sutarto 132 Surakarta Kode pos 57126 Telp (0271) 634 634,
Faksimile (0271) 637412 Email :
Website : rsmoewardi.jatengprov.go.id

Surakarta, 01 Desember 2015

Nomor : 1092 / DIK / XII / 2015
Lampiran : -
Perihal : Pengantar Penelitian

Kepada Yth. :

1. Ka. Instalasi Rekam Medik
2. Ka. Ruang Poliklinik Penyakit Dalam

RSUD Dr. Moewardi
di-

SURAKARTA

Memperhatikan Surat dari Ka. Prodi Informatika Fak Komunikasi & Informatika UMS Surakarta Nomor :103/C.6-II.3/Inf-FKI/XI/2015; perihal Permohonan Ijin penelitian dan disposisi Direktur tanggal 16 November 2015, maka dengan ini kami menghadapkan mahasiswa:

Nama : Taufiq Usman

NIM : L 200 120 054

Institusi : Prodi S.1 Informatika Fak. Komunikasi & Informatika UMS


Untuk melaksanakan penelitian dalam rangka pembuatan **Skripsi** dengan judul : "**Perbandingan 3 Metode Data Mining untuk Menentukan Hipertensi di Rumah Sakit Umum Daerah Dr. Moewardi**".

Demikian untuk menjadikan periksa dan atas kerjasamanya diucapkan terima kasih.

Kepala
Bagian Pendidikan & Penelitian,


dr. Eva Niamazilawati, Sp.PD, M.Kes

SIP : 33724.57126/DS/01/449.1/0066/03/2011


Slamet Gunanto, SKM. M.Kes
NIP. 19660310 198902 1 002

Tembusan Kepada Yth.:

1. Wadir Umum RSDM (sebagai laporan)
2. Ka. Inst. Rajal
3. Arsip

RSDM Cepat, Tepat, Nyaman dan Mudah



**KOMISI ETIK PENELITIAN KESEHATAN
RSUD Dr. Moewardi
Fakultas Kedokteran Universitas sebelas Maret**



BUKTI PENGAJUAN KELAIKAN ETIK

Yang bertanda tangan dibawah ini menyatakan bahwa data yang saya isikan adalah benar.

Peneliti : TAUFIQ USMAN
: L200120054
Judul Penelitian : PERBANDINGAN 3 METODE DATA MINING UNTUK MENENTUKAN
: HIPERTENSI DI RUMAH SAKIT UMUM DAERAH dr, MOEWARDI
SURAKARTA
Lokasi Tempat Penelitian : Poli Klinik Dalam



L200120054 - 2363

Mengetahui
Petugas


Surakarta : 19 Nov 2015

Peneliti



(TAUFIQ USMAN)

L200120054



HEALTH RESEARCH ETHICS COMMITTEE
KOMISI ETIK PENELITIAN KESEHATAN

Dr. Moewardi General Hospital
RSUD Dr. Moewardi

School of Medicine Sebelas Maret University
Fakultas Kedokteran Universitas Sebelas Maret



ETHICAL CLEARANCE
KELAIKAN ETIK

Nomor : 809 / XI / HREC / 2015

The Health Research Ethics Committee Dr. Moewardi General Hospital / School of Medicine Sebelas

Komisi Etik Penelitian Kesehatan RSUD Dr. Moewardi / Fakultas Kedokteran Universitas Sebelas Maret

Maret University Of Surakarta, after reviewing the proposal design, herewith to certify
Surakarta, setelah menilai rancangan penelitian yang diusulkan, dengan ini menyatakan

That the research proposal with topic :
Bahwa usulan penelitian dengan judul

PERBANDINGAN 3 METODE DATA MINING UNTUK MENENTUKAN HIPERTENSI
DI RUMAH SAKIT UMUM DAERAH Dr. MOEWARDI SURAKARTA

Principal investigator : TAUFIQ USMAN
Peneliti Utama L200120054

Location of research : RSUD Dr. Moewardi Surakarta
Lokasi Tempat Penelitian

Is ethically approved
Dinyatakan laik etik



Issued on : 28 November 2015

Chairman
Ketua

Dr. Hari Wujoso, dr., Sp.F.MM
NIP. 19621022 199503 1 001